Distinguishing $\text{bla}_{\text{KPC}}$ Gene-Containing IncF Plasmids from Epidemiologically Related and Unrelated Enterobacteriaceae Based on Short- and Long-Read Sequence Data

Joep J. M. Stohr, a,b Marjolein F. Q. Kluytmans-van den Bergh, a,c,d Veronica A. T. C. Weterings, a John W. A. Rossen, e,f Jan A. J. W. Kluytmans a,b,c,d

aDepartment of Infection Control, Amphia Hospital, Breda, the Netherlands
bLaboratory for Medical Microbiology and Immunology, Elisabeth-TweeSteden Hospital, Tilburg, the Netherlands
cAmphia Academy Infectious Disease Foundation, Amphia Hospital, Breda, the Netherlands
dJulius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
eDepartment of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
fDepartment of Pathology, University of Utah, School of Medicine, Salt Lake City, Utah, USA

ABSTRACT Limited information is available on whether $\text{bla}_{\text{KPC}}$-containing plasmids from isolates in a hospital outbreak can be differentiated from epidemiologically unrelated $\text{bla}_{\text{KPC}}$-containing plasmids based on sequence data. This study aimed to evaluate the performance of three approaches to distinguish epidemiologically related from unrelated $\text{bla}_{\text{KPC}}$-containing pKpQIL-like IncFII(k2)-IncFIB(pQIL) plasmids. Epidemiologically related isolates were subjected to short- and long-read whole-genome sequencing. A hybrid assembly was performed, and plasmid sequences were extracted from the assembly graph. Epidemiologically unrelated plasmid sequences were extracted from GenBank. Pairwise comparisons of epidemiologically related and unrelated plasmids based on SNPs using snippy and of phylogenetic distance using Roary and using a similarity index that penalizes size differences between plasmids (Stoesser index) were performed. The percentage of pairwise comparisons misclassified as genetically related or as clonally unrelated was determined using different genetic thresholds for genetic relatedness. The ranges of number of SNPs, Roary phylogenetic distance, and Stoesser index overlapped between the epidemiologically related and unrelated plasmids. When a genetic similarity threshold that classified 100% of epidemiologically related plasmid pairs as genetically related was used, the percentages of plasmids misclassified as epidemiologically related ranged from 6.7% (Roary) to 20.8% (Stoesser index). Although epidemiologically related plasmids can be distinguished from unrelated plasmids based on genetic differences, $\text{bla}_{\text{KPC}}$-containing IncF plasmids from epidemiologically related and unrelated Enterobacteriaceae based on short- and long-read sequence data. Antimicrob Agents Chemother 65:e00147-21. https://doi.org/10.1128/AAC.00147-21.

Copyright © 2021 American Society for Microbiology. All Rights Reserved.
Address correspondence to Joep J. M. Stohr, joep.stohr@gmail.com.
Received 27 January 2021
Returned for modification 8 March 2021
Accepted 26 March 2021
Accepted manuscript posted online 5 April 2021
Published 18 May 2021

Infections with carbapenem-resistant Enterobacteriaceae are associated with increased mortality and have emerged as an urgent public health threat (1–3). Worldwide, one of the most frequent mechanisms for carbapenem resistance in Enterobacteriaceae is the production of Klebsiella pneumoniae carbapenemase (KPC) (4). Nosocomial transmission plays an important role in spreading these KPC-producing Enterobacteriaceae, and several hospital outbreaks have been described (5–8). Molecular typing is often used to detect the source and route of transmission in these hospital outbreak settings.
Whole-genome sequencing of bacterial isolates is currently the ultimate tool for molecular typing, enabling comparison of the entire bacterial chromosome to identify bacterial clones with great precision using either a gene-by-gene or single-nucleotide-polymorphism (SNP) approach (10). However, the gene encoding KPC production, \( \text{bla}_{\text{KPC}} \), is located not on the bacterial chromosome but on large conjugative resistance plasmids. These plasmids can transfer between isolates, and nosocomial transmission of these plasmids can go undetected when only molecular typing of the bacterial chromosome is performed (11).

One of the most predominant plasmid types containing \( \text{bla}_{\text{KPC}} \) are the pKpQiL-like IncFII(k2)-IncFIB(pQiL) plasmids, which have spread across the globe (12–14). Several reports have already described plasmid spread between different isolates and patients during hospital outbreaks (15–19). Analyzing these resistance plasmids is typically performed by generating a combination of short- and long-read sequence data, enabling hybrid assembly algorithms to perform complete plasmid assemblies (17, 20–22). Accurately distinguishing epidemiologically related from unrelated plasmids (e.g., based on the number of SNPs) is essential to detect or exclude nosocomial plasmid transmission in outbreaks. However, limited information is available on whether \( \text{bla}_{\text{KPC}} \)-containing plasmids from isolates in a hospital outbreak can be differentiated from epidemiologically unrelated \( \text{bla}_{\text{KPC}} \)-containing plasmids based on sequence data. This study aimed to evaluate the performance of three approaches, based on determining SNPs, phylogenetic distance, and a similarity index, to distinguish epidemiologically related from unrelated \( \text{bla}_{\text{KPC}} \)-containing pKpQiL-like IncFII(k2)-IncFIB(pQiL) plasmids.

RESULTS

The percentage of whole-genome multilocus sequence typing (wgMLST) allele differences between the 3 \textit{K. pneumoniae} isolates ranged from 0.002% (Pk1 versus Pk2) to 0.770% (Pk1 and Pk2 versus Pk4). The percentage of wgMLST allele differences between the 2 \textit{Klebsiella aerogenes} isolates was 0.979%. In all epidemiologically related isolates, a circular plasmid contig containing a \( \text{bla}_{\text{KPC-2}} \) gene (located within a Tn4401a transposon) and an IncFII(k2) and IncFIB plasmid replicon was detected and extracted from the assembly graph (Table 1). The plasmid contigs were either 113,638 or 113,639 bp, had a GC content of 53.9%, and contained the acquired beta-lactam resistance genes \( \text{bla}_{\text{TEM1A}} \) and \( \text{bla}_{\text{OXA-9}} \). Fifteen epidemiologically unrelated plasmids were detected in and extracted from GenBank. The plasmids were isolated from patients in 5 different countries (Greece, Italy, the United States, Australia, and the United Kingdom) between 2007 and 2017 and ranged in size from 99,142 to 117,916 bp (Table 1) (13, 14, 23). Global sequence alignment revealed 13 plasmids to be entirely homologous (see Fig. S1 in the supplemental material). In the other 8 plasmids, only a limited number of structural rearrangements, deletions, or/and insertions were detected (Fig. S1).

The number of pairwise comparisons was 15 for the epidemiologically related plasmids and 120 for the epidemiologically unrelated plasmids (including the plasmid from the index patient) (Table 2). The number of aligned bases used for the SNP analysis between each plasmid and the reference ranged from 80,591 to 111,332 (see Table S1). The median number of SNPs varied significantly between the two groups (\( P < 0.001 \)) (Table 2). However, the range of SNPs overlapped, ranging from 0 to 1 for the epidemiologically related plasmids and from 0 to 674 for the epidemiologically unrelated plasmids (Table 2; Table S2). A total of 192 genes were detected, of which 56 were present in all plasmids. The phylogenetic distance as determined using Roary ranged from 0.00 to 0.06 between the epidemiologically related plasmids and from 0.00 to 1.69 between the epidemiologically unrelated plasmids (\( P < 0.001 \)) (Table 2; Table S3). The Stoesser similarity index ranged from 0.99 to 1 between the epidemiologically related plasmids and from 0.51 to 1 between the epidemiologically unrelated plasmids (\( P < 0.001 \)) (Table 2; Table S4). Between the three comparison methods, the number of SNPs and the Roary phylogenetic distance showed the highest degree of
correlation, with a Spearman’s rank correlation coefficient of 0.820 ($P < 0.001$). The Spearman’s rank correlation coefficients between the Stoesser similarity index and the number of SNPs or Roary phylogenetic distance were 0.65 ($P < 0.001$) and 0.77 ($P < 0.001$), respectively.

When the threshold was set at the minimal value (or maximal value for the Stoesser similarity index) that classified 100% of epidemiologically related plasmid pairs as genetically related, the percentage of presumed epidemiologically unrelated plasmid pairs that were classified as epidemiologically related was 12.5% with the SNPs approach (Fig. 1A), 6.7% with the Roary phylogenetic distance approach (Fig. 1B), and 20.8% for the Stoesser similarity index approach (Fig. 1C).

**DISCUSSION**

In this study on the genetic diversity of blaKPC-containing IncF plasmids, epidemiologically related plasmids were on average genetically more related than epidemiologically unrelated plasmids. However, the genetic diversity of plasmids was limited, which

### TABLE 2 Number of pairwise comparisons, SNPs, genes variably present or present, and combined number of SNPs and variable gene presence between the plasmids

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SNPs</th>
<th>Phylogenetic distance using Roary</th>
<th>Stoesser similarity index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Related</td>
<td>Unrelated</td>
<td>Related</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>29</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>0–1</td>
<td>0–674</td>
<td>0.00–0.06</td>
</tr>
</tbody>
</table>

*The number of pairwise comparisons for related plasmids was 15, and the number for unrelated plasmids was 120.*
precluded 100% correct classification of epidemiologically related and unrelated isolates. When a genetic similarity threshold that classified 100% of epidemiologically related plasmid pairs as genetically related was used, the percentages of plasmids misclassified as epidemiologically related based on the sequence data ranged from 6.7% to 20.8% depending on the comparison method. The phylogenetic distance determined using Roary showed the highest discriminatory power between the epidemiologically related and unrelated plasmids. Since Roary relies primarily on differences in gene presence to determine the phylogenetic distance between sequences, solely using Roary to differentiate epidemiologically related from unrelated plasmid sequences could lead to falsely dismissing plasmid transmission when large rearrangements occur between different plasmids present in an outbreak. These within-outbreak plasmid rearrangements have been described in previous studies (18, 24). Within-outbreak rearrangements might have less effect on a SNP-based analysis, which uses only blocks of shared sequences for its analysis. Therefore, it might be necessary to tailor the method used, e.g., SNPs or Roary (or both), to differentiate between epidemiologically related and unrelated plasmids to each specific plasmid outbreak setting. The relatively poor performance of the Stoesser index in this analysis could be due to the presence of 9 plasmids of equal sequence length (5 epidemiologically related; 4 epidemiologically unrelated), which is an important parameter in determining the phylogenetic distance in this method.

Molecular plasmid typing has been used to demonstrate the transmission of blaKPC-containing plasmids between bacterial isolates and patients (16, 17). However, previous studies did not include similar plasmids of unrelated patients in the analysis. Our findings on the degree of sequence similarity between blaKPC-containing plasmids isolated from epidemiologically related and unrelated patients can be used to reveal a possible horizontal transfer of blaKPC-containing IncFII(k2)-IncFIB(pQiL) plasmids in

FIG 1 Percentage of pairwise comparisons of epidemiologically related and unrelated plasmids misclassified as either genetically related (blue) or genetically unrelated (red) using different thresholds of (A) number of SNPs, (B) Roary phylogenetic distance, and (C) Stoesser similarity index. Green dotted lines with numbers in boxes show minimal threshold values (or maximal value for the Stoesser similarity index) that classified 100% of epidemiologically related plasmid pairs as genetically related.
hospital outbreaks. Despite this, when the lowest (or highest for the Stoeser-similarity index) threshold for clonally relatedness that classified 100% of epidemiologically related plasmid pairs as genetically related was used, the percentage of presumed epidemiologically unrelated plasmid pairs that were classified as genetically related was higher in blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmids than the percentages described when the bacterial chromosome was typed in a previous study (25). This relatively high percentage of misclassifications compared to bacterial chromosome typing was present in all three comparison methods, suggesting a relatively stable plasmid content. Several other studies have also described a high degree of sequence similarity between blaKPC-containing plasmids isolated from epidemiologically unrelated patients (13, 23). However, to the best of our knowledge, this is the first study comparing the sequence similarity between epidemiologically related and unrelated blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmids.

The limited number of differences between plasmids isolated from epidemiologically unrelated patients observed in this study is also seen for other resistance plasmids (26, 27). OXA-48-containing plasmids belonging to the IncL/M replicon type have a far wider host range than the blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmids included in our study. Despite this, a recent study has stated that even these OXA-48 containing plasmids from outbreak-related patients could not be distinguished from similar plasmids of non-outbreak-related patients based on SNPs (26). Interestingly, other studies found blaKPC-containing plasmids to be highly variable in their nucleotide sequence within the same outbreak and even within patients (18, 28). This suggests that the plasmid content could also be highly unstable in vivo during hospital outbreaks. Therefore, it could well be that transmission of blaKPC-containing IncF plasmids within hospital outbreaks cannot be dismissed based on sequence dissimilarity between the different plasmids investigated.

This study included bacterial isolates encompassing three different species isolated from an extensively described outbreak with clear epidemiological links between the different outbreak patients. Moreover, the epidemiologically related blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmids were compared to a set of similar unrelated plasmids using both a SNP and gene-by-gene approach in a setting where blaKPC-containing Enterobacteriaceae are nonendemic. This limits the possibility of falsely classifying blaKPC-containing plasmids as epidemiologically related. Furthermore, all epidemiologically related blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmid sequences in this study were single circular contigs in the assembly graph.

This study has several limitations. Only blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmid sequences of six isolates in one outbreak were included. Therefore, it remains unknown whether plasmids belonging to other incompatibility groups and containing other resistance genes also show a high degree of sequence similarity to epidemiologically related plasmids and whether Roary also shows the highest discriminatory power between the epidemiologically related and unrelated plasmids in other plasmid types. Moreover, studies that include more epidemiologically related and unrelated blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmids are needed to confirm our findings. Additionally, the relatively high occurrence of horizontal gene transfer of blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmids detected in the outbreak included here could be facilitated by the use of antibiotics in 3 of the 6 patients included in our outbreak. Therefore, future studies should be performed, including epidemiologically related blaKPC-containing IncFII(k2)-IncFIB(pQIL) plasmids isolated in other outbreaks, to confirm our findings.

To conclude, although epidemiologically related blaKPC-containing IncF plasmids can be distinguished from unrelated blaKPC-containing IncF plasmids based on genetic differences, epidemiologically related and unrelated blaKPC-containing IncF plasmids show a high degree of sequence similarity. The phylogenetic distance as determined using Roary showed the highest degree of discriminatory power between the epidemiologically related and unrelated blaKPC-containing IncF plasmids.
**MATERIALS AND METHODS**

Outbreak and plasmid selection. From June to December 2013, an outbreak occurred in an 800-bed teaching hospital (approximately 40,000 admissions/year) and a 130-bed nursing home in Breda, the Netherlands. In total, six patients were colonized or infected with KPC-producing Enterobacteriaceae belonging to three different species (K. pneumoniae, K. aerogenes, and Escherichia coli). The outbreak was recognized on 13 July 2013, when a KPC-producing K. pneumoniae isolate was detected in the pleural fluid of patient 2 in the pulmonary ward. This event followed the earlier repatriation of patient 1 from an intensive care unit in Greece on June 2013, who was found to be colonized with a KPC-producing K. pneumoniae isolate 2 days after admission. Patient 1 was transferred to a nursing home in August 2013, where further transmission occurred between November and December 2013 from patient 1 to 4 other patients in the nursing home. The outbreak was comprehensively described by Weterings et al. (6).

Epidemiologically related isolates (i.e., belonging to the outbreak) were selected in such a way that at least one unique 
\[ \text{bla}_{	ext{KPC}} \] containing isolate, based on species identification, antimicrobial susceptibility testing, and molecular typing (6), was included. Moreover, one additional KPC-producing K. pneumoniae isolate was included that was cultured only 2 months after the outbreak period in February 2014 from a rectal swab taken from patient 3. The antimicrobial susceptibility testing results for this isolate differed from those of the outbreak KPC-producing K. pneumoniae isolates, as it was susceptible to ciprofloxacin, tobramycin, and trimethoprim-sulfamethoxazole according to Vitek-2 (bioMérieux, Marcyl’Étoile, France) automated susceptibility testing using EUCAST breakpoints (v10.0). Epidemiologically unrelated plasmid sequences containing a 
\[ \text{bla}_{	ext{KPC}} \] gene and the replicon IncFIB, FIB were extracted from GenBank on 11 September 2017 (see the supplemental materials and methods). To ensure epidemiological unrelatedness, only the first uploaded plasmid sequence per geographic location per year was included, and no sequences with a Dutch origin and isolated in 2013/2014 were included in the analysis.

Whole-genome sequencing and analysis. Epidemiologically related isolates were subjected to short-read sequencing on an Illumina MiSeq platform using Nextera XT chemistry generating 250-bp paired-end reads (Illumina, San Diego, CA, USA) and long-read sequencing on a MinION sequencer using the FLO-MIN106D flow cell and the SQK RKBN004 rapid barcoding sequencing kit according to the standard protocol provided by the manufacturer (Oxford Nanopore Technologies, Oxford, United Kingdom). A hybrid assembly of long-read and short-read sequence data was performed using Unicycler v.0.8.4 (29). Whole-genome MLST (wgMLST) (core and accessory genome) was performed for all sequenced isolates using Ridom SeqSphere+, version 4.1.9 (Ridom, Münster, Germany). Species-specific typing schemes were used as described by Kluytmans-van den Bergh et al. (25). The pairwise genetic difference between isolates of the same species was calculated by dividing the number of allele differences by the total number of alleles shared between the two sequences in the wgMLST typing scheme. The sequenced isolates’ genomes were uploaded to the online bioinformatic tools ResFinder v.3.1 and PlasmidFinder v.2.0 (Center for Genomic Epidemiology, Technical University of Denmark, Lyngby, Denmark) (30, 31). Acquired resistance genes were called when at least 60% of the length of the best matching gene in the ResFinder database was covered with a sequence identity of at least 90%. Plasmid replicon genes were called when at least 60% of the sequence length of the replicon gene in the PlasmidFinder database was covered with a sequence identity of at least 80%.

Plasmid analysis. Circular components created by the hybrid assembly that were smaller than 1,000 kb and contained a 
\[ \text{bla}_{	ext{KPC}} \] gene were extracted from the assembly graph using BANDAGE v0.8.1 (32). All extracted plasmid components and the plasmid sequences extracted from the GenBank were annotated using Prokka v1.13.3 (33). A global sequence alignment of the extracted plasmids was performed using progressiveMAUVE v2.4.0. The number of SNPs between the plasmids were calculated using snippy v4.4.5 (reference sequence: GenBank accession number NC_014016.1) (32). A pangenome was constructed, and phylogenetic distance based on gene presence/absence between the different plasmids was determined using Roary v1.13.2 (34). All plasmids were pairwise aligned using dnadiff, and a similarity index was calculated between the different plasmids as described by Stoesser et al. (28) (Stoesser similarity index). This similarity index can vary from 0 (completely unsimilar plasmids) to 1 (identical plasmids) and, contrary to both Roary phylogenetic distance and number of SNPs, penalizes size differences between plasmids. Pairwise comparisons of SNPs, Roary phylogenetic distance, and Stoesser similarity were performed between (i) epidemiologically related plasmids and (ii) the first plasmid isolated from the index patient in the outbreak and the epidemiologically unrelated plasmid sequences extracted from GenBank. The percentage of pairwise comparisons misclassified as genetically related or as clonally unrelated was determined using different genetic thresholds for genetic relatedness for all three comparison methods: SNP thresholds tested ranged from 0 to 50 with steps of 1, Roary thresholds from 0.0 to 0.5 with steps of 0.01, and Stoesser index thresholds from 0.5 to 1 with steps of 0.0001.

Statistical analysis. The median number of SNPs, the median Roary phylogenetic distance, and the median Stoesser similarity index were calculated for epidemiologically related and unrelated plasmids. Medians were compared using Mann-Whitney U tests (SciPy v1.5.0). Spearman’s rank correlation coefficients between the number of SNP differences, Roary phylogenetic distance, and the Stoesser similarity index were calculated using SciPy v1.5.0.

Data availability. Short- and long-read sequence data included in this study are available from the European Nucleotide Archive of the European Bioinformatics Institute under study accession number PRJEB41009.

June 2021 Volume 65 Issue 6 e00147-21 aac.asm.org
SUPPLEMENTAL MATERIAL
Supplemental material is available online only.
SUPPLEMENTAL FILE 1, XLSX file, 0.01 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.01 MB.
SUPPLEMENTAL FILE 3, XLSX file, 0.01 MB.
SUPPLEMENTAL FILE 4, XLSX file, 0.01 MB.
SUPPLEMENTAL FILE 5, PDF file, 1.7 MB.

ACKNOWLEDGMENT
We declare no competing interests.

REFERENCES


